NASA
Technical
Paper
2529

May 1986

NASA-TP-2529 19860019189

Solution of Elliptic Partial Differential Equations by Fast Poisson Solvers Using a Local Relaxation Factor

I—One-Step Method

Sin-Chung Chang

NASA

NASA Technical Paper 2529

1986

Solution of Elliptic Partial Differential Equations by Fast Poisson Solvers Using a Local Relaxation Factor

I—One-Step Method

Sin-Chung Chang

Lewis Research Center
Cleveland, Ohio

NASA
National Aeronautics
and Space Administration

Scientific and Technical Information Branch

Summary

An algorithm for solving a large class of two- and three-dimensional nonseparable elliptic partial differential equations (PDE's) is developed and tested. It uses a modified D'Yakanov-Gunn iterative procedure in which the relaxation factor is grid-point dependent. It is easy to implement and applicable to a variety of boundary conditions. It is also computationally efficient, as indicated by the results of numerical comparisons with other established methods. Furthermore the current algorithm has the advantage of possessing two important properties which the traditional iterative methods lack; that is, (1) the convergence rate is relatively insensitive to grid-cell size and aspect ratio, and (2) the convergence rate can be easily estimated by using the coefficient of the PDE being solved.

Introduction

Since the middle sixties, fast direct solvers (FDS's) have been developed for the numerical solution of separable elliptic partial differential equations (PDE's) (refs. 1 to 5). Based on Fourier analysis and cyclic reduction, FDS algorithms are most effective on a uniform rectangular grid. They can obtain the solution with efficiency far beyond the reach of traditional iterative procedures such as successive overrelaxation (SOR) methods.

Generally, FDS algorithms are not directly applicable to an elliptic problem with either a computation domain of irregular shape or a nonseparable PDE. The limitation of the computation domain may be circumvented either by mapping the original domain onto a rectangular domain or by using the capacity matrix method (ref. 5). The limitation of a nonseparable PDE can be circumvented by a semidirect procedure, that is, an iterative procedure driven by an FDS. In this study, a new semidirect procedure is developed and used as an elliptic solver for both two-dimensional (2-D) and three-dimensional (3-D) problems. This new iterative procedure is easy to implement, computationally efficient, and applicable to a variety of boundary conditions. Furthermore it has the advantage of possessing two important properties which the traditional iterative methods lack; that is, (1) the convergence rate is insensitive to grid-cell size and aspect ratio, and (2) the convergence rate can be easily estimated with the coefficients of the PDE being solved.

Many elliptic PDE's can be expressed as

$$Qu = h \tag{1}$$

where Q is a nonseparable second-order linear elliptic operator, u the dependent variable, and h a given source term. Equation (1) may be solved with the iterative procedure

$$P(u^{n+1} - u^n) = -\tau(Qu^n - h) \tag{2}$$

where n is the iteration number, τ a nonzero relaxation factor, and P a separable elliptic operator, which can be directly inverted by an FDS. This procedure is a continuous analogue of the D'Yakanov-Gunn iterations (ref. 6) and was utilized by Concus and Golub (ref. 7) and Bank (ref. 8) in their works on the numerical solution of nonseparable elliptic equations. In the previous works involving iteration (2) the relaxation factor τ is treated as a constant and the iteration is accelerated by an optimal choice of τ. In the current report, a more efficient algorithm is obtained by using a spatially varying relaxation factor.

The use of a local (spatially varying) relaxation factor in the current study is motivated by an earlier study of a semidirect procedure (ref. 9). In the previous study, the local convergence rate evaluated by using a simple von Neumann analysis, to a great extent, is consistent with the numerical results. Based on this observation, it becomes obvious that a local relaxation factor could be used in iteration (2) to optimize its local convergence rate. Recently, a similar idea was also used by Botta and Veldman (ref. 10) to develop their SOR-related local relaxation method. However, as shown later, there is an underlying reason which makes the use of a local relaxation factor in the current procedure particularly attractive.

As shown by the work of Bank (ref. 8), iteration (2) can also be accelerated by choosing an operator P, which closely resembles the operator Q. Application of this technique, however, could be limited by the following considerations:

(1) This technique may require the use of a general separable operator P. This, however, is computationally inefficient, since an FDS code for a general separable operator is about five times slower than one for the Laplacian ∇^2 (ref. 5).

(2) To apply this technique, the FDS code for the operator P, generally, must be made to individual specifications. This may require a considerable effort.

The preceding considerations lead us to choose $P = \nabla^2$ or its equivalent in the current study.

In the section Analysis, the convergence rate of the central difference form of equation (2) is studied for a constant coefficient operator Q by assuming the iterative errors satisfy the periodic boundary conditions. The analysis is a rigorous version of the von Neumann analysis and its results are used to determine the optimal value of the relaxation factor τ. In the section Local Relaxation, the results obtained in the section Analysis are extended to solve PDE's with variable coefficients. In the section Numerical Evaluation, the current method is numerically evaluated with a variety of 2-D nonseparable elliptic PDE's. In addition to the advantages noted previously, the results of this numerical evaluation indicate that the current procedure can be used to solve PDE's with a cross-derivative term and that it works very well for many PDE's with rapidly varying coefficients.

Finally, in the section Application to a 3-D Flow Problem, the current procedure is incorporated into an Euler Solver (ref. 9) to obtain solutions for 3-D incompressible flows in a 180° turning channel. It is shown that this new procedure converges with a rate much higher than that reported in reference 9. The successive line overrelaxation (SLOR) calculations referred to in the section Numerical Evaluation were carried out by using a code developed by Shih-Hung Chang of Cleveland State University.

Analysis

As an initial step, iteration (2) is studied by assuming that τ is a constant and

$$Q \overset{\text{def.}}{=} a\frac{\partial^2}{\partial x^2} + 2b\frac{\partial^2}{\partial x \partial y} + c\frac{\partial^2}{\partial y^2} \tag{3}$$

where a, b, and c are arbitrary constants subjected to the elliptic conditions

$$\left.\begin{aligned} a &> 0 \\ c &> 0 \\ ac - b^2 &> 0 \end{aligned}\right\} \tag{4}$$

Furthermore, it is assumed that

$$P \overset{\text{def.}}{=} a_o\frac{\partial^2}{\partial x^2} + c_o\frac{\partial^2}{\partial y^2} \qquad (a_o > 0,\ c_o > 0) \tag{5}$$

where a_o and c_o are two arbitrary positive constants. When a uniform grid with grid intervals Δx and Δy in the x- and the y-directions is used, the central difference forms of equations (1) and (2) at a grid point (i,j) are

$$\widetilde{Q}(u_{i,j}) = h_{i,j} \tag{6}$$

and

$$\widetilde{P}\left(u_{i,j}^{n+1} - u_{i,j}^{n}\right) = -\tau\left[\widetilde{Q}\left(u_{i,j}^{n}\right) - h_{i,j}\right] \tag{7}$$

respectively. Here $h_{i,j}$ is the source term and the finite difference operators $\widetilde{Q}$ and $\widetilde{P}$, respectively, are defined by

$$\begin{aligned} \widetilde{Q}(v_{i,j}) \overset{\text{def.}}{=} & \ a(\Delta x)^{-2}(v_{i+1,j} + v_{i-1,j} - 2v_{i,j}) \\ & + c(\Delta y)^{-2}(v_{i,j+1} + v_{i,j-1} - 2v_{i,j}) \\ & + b(2\Delta x \Delta y)^{-1}(v_{i+1,j+1} + v_{i-1,j-1} \\ & - v_{i+1,j-1} - v_{i-1,j+1}) \end{aligned} \tag{8}$$

and

$$\begin{aligned} \widetilde{P}(v_{i,j}) \overset{\text{def.}}{=} & \ a_o(\Delta x)^{-2}(v_{i+1,j} + v_{i-1,j} - 2v_{i,j}) \\ & + c_o(\Delta y)^{-2}(v_{i,j+1} + v_{i,j-1} - 2v_{i,j}) \end{aligned} \tag{9}$$

where $v_{i,j}$ is any function of the grid point (i,j). The operator $\widetilde{P}$ can be considered as a central difference Poisson operator for a uniform grid with grid intervals $\Delta x/\sqrt{a_o}$ and $\Delta y/\sqrt{c_o}$. Thus, it may be inverted by using a fast Poisson solver.

To study the convergence rate of iterative procedure (7), one notes that equations (6) and (7) imply that

$$\widetilde{P}\left(e_{i,j}^{n+1} - e_{i,j}^{n}\right) = -\tau\widetilde{Q}\left(e_{i,j}^{n}\right) \tag{10}$$

where

$$e_{i,j}^{n} \overset{\text{def.}}{=} u_{i,j}^{n} - u_{i,j} \tag{11}$$

Given a set of $e_{i,j}^{0}$'s which satisfies the periodic conditions

$$e_{i,j}^{0} = e_{i+Kj}^{0} = e_{i,j+L}^{0} \qquad (i,j = 0, \pm 1, \pm 2, \ldots) \tag{12}$$

where $K(\geq 2)$ and $L(\geq 2)$ are two arbitrary integers, it is shown in appendix A that

(1) $e_{i,j}^{n}$'s $(n = 1,2,\ldots;\ i,j = 0, \pm 1, \pm 2, \ldots)$ are uniquely determined by equation (10) and the following auxiliary conditions:

$$e_{i,j}^{n} = e_{i+Kj}^{n} = e_{i,j+L}^{n} \qquad (n = 1,2,\ldots) \tag{13}$$

and

$$\sum_{i=0}^{(K-1)} \sum_{j=0}^{(L-1)} e_{i,j}^{n} = 0 \qquad (n = 1,2,\ldots) \tag{14}$$

(2) Let

$$\varphi_{i,j}^{(k,\ell)} \overset{\text{def.}}{=} \frac{1}{\sqrt{KL}} \exp\left[2\pi I\left(\frac{k \bullet i}{K} + \frac{\ell \bullet j}{L}\right)\right] \qquad I \equiv \sqrt{-1}$$

$$(i,j = 0, \pm 1, \pm 2, \ldots;\ k = 0,1,2,\ldots,(K-1);$$

$$\ell = 0,1,2,\ldots,(L-1)) \tag{15}$$

$$\sigma_p^{(k,\ell)} \overset{\text{def.}}{=} -4\left\{a_o\left[\frac{1}{\Delta x}\sin(\pi k/K)\right]^2 + c_o\left[\frac{1}{\Delta y}\sin(\pi \ell/L)\right]^2\right\}$$

$$(k = 0,1,2,\ldots,(K-1);\ \ell = 0,1,2,\ldots,(L-1)) \tag{16}$$

$$\sigma_q^{(k,\ell)} \overset{\text{def.}}{=} -4\left\{a\left[\frac{1}{\Delta x}\sin(\pi k/K)\right]^2 + c\left[\frac{1}{\Delta y}\sin(\pi \ell/L)\right]^2 + 2b\left[\frac{1}{\Delta x}\sin(\pi k/K)\right]\left[\frac{1}{\Delta y}\sin(\pi \ell/L)\right] \times \cos(\pi k/K)\cos(\pi \ell/L)\right\}$$

$$(k = 0,1,2,\ldots,(K-1);\ \ell = 0,1,2,\ldots,(L-1)) \tag{17}$$

$$E^{0,(k,\ell)} \overset{\text{def.}}{=} \sum_{i=0}^{(K-1)} \sum_{j=0}^{(L-1)} e_{i,j}^0 \overline{\varphi_{i,j}^{(k,\ell)}} \qquad (k,\ell) \in \Psi \tag{18}$$

and

$$G^{(k,\ell)}(\tau) \overset{\text{def.}}{=} 1 - \tau\left(\frac{\sigma_q^{(k,\ell)}}{\sigma_p^{(k,\ell)}}\right) \qquad (k,\ell) \in \Psi \tag{19}$$

where $\in$ means an element of, and Ψ is the set of ordered pairs defined by

$$\Psi \overset{\text{def.}}{=} \{(k,\ell) \mid k = 0,1,2,\ldots,(K-1);$$

$$\ell = 0,1,2,\ldots,(L-1);\ (k,\ell) \neq (0,0)\} \tag{20}$$

Then the unique solution to equations (10), (13), and (14) is explicitly given by ($n = 1,2,\ldots$)

$$e_{i,j}^n = \sum_{(k,\ell)\in\Psi} \left[G^{(k,\ell)}(\tau)\right]^n \bullet E^{0,(k,\ell)} \bullet \varphi_{i,j}^{(k,\ell)} \tag{21}$$

One notes that $G^{(k,\ell)}(\tau)$ is well defined for all $(k,\ell) \in \Psi$, since

$$\sigma_p^{(k,\ell)} < 0 \qquad \text{if } (k,\ell) \in \Psi \tag{22}$$

This inequality follows from the assumptions $a_o > 0$ and $c_o > 0$, and the fact that (0,0) does not belong to Ψ.

As defined in equation (11), $e_{i,j}^n$ is the error of nth iterative solution. According to equation (21), this error is a sum of $(K \times L - 1)$ terms, and each term is multiplied by the factor $G^{(k,\ell)}(\tau)$ as the iteration number n increases by 1. Obviously, the term with the greatest value of $|G^{(k,\ell)}(\tau)|$ eventually becomes dominant if the corresponding $E^{0,(k,\ell)}$ does not vanish. Let the error norm $\|e^n\|$ and the asymptotic error multiplication factor M^∞, respectively, be defined as

$$\|e^n\| \overset{\text{def.}}{=} \left[\sum_{i=0}^{(K-1)} \sum_{j=0}^{(L-1)} \left(e_{i,j}^n\right)^2\right]^{1/2} \tag{23}$$

and

$$M^\infty \overset{\text{def.}}{=} \lim_{n\to+\infty} \frac{\|e^{n+1}\|}{\|e^n\|} \tag{24}$$

Then, assuming every $E^{0,(k,\ell)} \neq 0$, it may be concluded that

$$\lim_{n\to+\infty} \frac{|e_{i,j}^{n+1}|}{|e_{i,j}^n|} = G(\tau) \qquad (i,j = 0, \pm 1, \pm 2, \ldots) \tag{25}$$

and

$$M^\infty = G(\tau) \tag{26}$$

where, for a given τ,

$$G(\tau) \overset{\text{def.}}{=} \operatorname*{Max}_{(k,\ell)\in\Psi} \left\{|G^{(k,\ell)}(\tau)|\right\}$$

A direct implication of equation (26) is that the value of M^∞ reaches its minimum if the parameter τ is chosen such that the function $G(\tau)$ is at its minimum. Let

$$\gamma^{(k,\ell)} \overset{\text{def.}}{=} \frac{\sigma_q^{(k,\ell)}}{\sigma_p^{(k,\ell)}}$$

and

$$\gamma_{\max} \overset{\text{def.}}{=} \operatorname*{Max}_{(k,\ell)\in\Psi} \{\gamma^{(k,\ell)}\}$$

$$\gamma_{\min} \overset{\text{def.}}{=} \operatorname*{Min}_{(k,\ell)\in\Psi} \{\gamma^{(k,\ell)}\}$$

It is shown in equation (33) that $\gamma_{\max} \geq \gamma_{\min} > 0$. As a result, one concludes from equation (19) that $G(\tau)$ reaches its minimum

$$G^o \overset{\text{def.}}{=} G(\tau^o) = \frac{\Sigma - 1}{\Sigma + 1} < 1 \tag{27}$$

when

$$\tau = \tau^o \overset{\text{def.}}{=} \frac{2}{\gamma_{\max} + \gamma_{\min}} \tag{28}$$

Here τ^o is the optimal relaxation factor, and

$$\Sigma \overset{\text{def.}}{=} \frac{\gamma_{\max}}{\gamma_{\min}} \tag{29}$$

Combining equations (26) and (27), one concludes that (1) $M^\infty < 1$ if $\tau = \tau^o$, and (2) M^∞ increases with an increase of Σ.

The values of $\gamma_{\max}$ and $\gamma_{\min}$, generally, are functions of the integers K and L. However, as will be shown, $\gamma_{\max}$ and $\gamma_{\min}$ approach two separate limits as the values of K and L increase. Let

$$\lambda_{\max} \overset{\text{def.}}{=} \frac{1}{2}\left(\hat{a} + \hat{c} + \sqrt{(\hat{a} - \hat{c})^2 + 4(\hat{b})^2}\right) \tag{30}$$

and

$$\lambda_{\min} \overset{\text{def.}}{=} \frac{1}{2}\left(\hat{a} + \hat{c} - \sqrt{(\hat{a} - \hat{c})^2 + 4(\hat{b})^2}\right) \tag{31}$$

where

$$\left.\begin{aligned} \hat{a} &\overset{\text{def.}}{=} \frac{a}{a_o} > 0 \\ \hat{c} &\overset{\text{def.}}{=} \frac{c}{c_o} > 0 \\ \hat{b} &\overset{\text{def.}}{=} \frac{b}{\sqrt{a_o c_o}} \end{aligned}\right\} \tag{32}$$

In appendix B, it is shown that

$$\lambda_{\max} \geq \gamma_{\max} \geq \gamma_{\min} \geq \lambda_{\min} > 0 \tag{33}$$

and

$$\left.\begin{aligned} \lim_{K,L \to +\infty} \gamma_{\max} &= \lambda_{\max} \\ \lim_{K,L \to +\infty} \gamma_{\min} &= \lambda_{\min} \end{aligned}\right\} \tag{34}$$

Thus, in the limit of $K,L \to +\infty$, τ^o and G^o, respectively, approach

$$\tau^* \overset{\text{def.}}{=} \frac{2}{\lambda_{\max} + \lambda_{\min}} \tag{35}$$

and

$$G^* \overset{\text{def.}}{=} \frac{\Sigma^* - 1}{\Sigma^* + 1} < 1 \tag{36}$$

where

$$\Sigma^* \overset{\text{def.}}{=} \frac{\lambda_{\max}}{\lambda_{\min}} \tag{37}$$

Two comments on equations (33) to (35) are as follows:

(1) The uniform bounds $\lambda_{\max}$ and $\lambda_{\min}$, generally do not exist if the operator $\widetilde{P}$ is replaced by an operator of other type.

(2) Since $\lambda_{\max} + \lambda_{\min} = \hat{a} + \hat{c}$,

$$\tau^* = \frac{2}{\hat{a} + \hat{c}} \tag{38}$$

Using equations (30) to (32), (36), and (37), it can be shown that the parameter G^* is a function of a, b, c, and c_o/a_o. If the coefficients a, b, and c are known, G^* becomes a function of the single variable c_o/a_o. As shown in appendix C, this function reaches its minimum

$$G^*_{\min} \overset{\text{def.}}{=} \frac{|b|}{\sqrt{ac}} \tag{39}$$

when

$$\frac{c_o}{a_o} = \frac{c}{a} \tag{40}$$

Furthermore, assuming $c_o/a_o = c/a$, it is shown in appendix C that

$$\tau^o = \tau^* \tag{41}$$

for any finite integers $K \geq 2$ and $L \geq 2$.

At this juncture, it is noted that equations (35) and (36) can also be derived (ref. 11), in a less rigorous fashion, by using a simple von Neumann analysis and theorem 1 in appendix B. The current analysis shows that the von Neumann analysis for equation (2.8) is justified only under many restricted

conditions. One of them, the uniqueness condition (14), generally is not required for other iterative procedures.

This section concludes with a discussion on the possible generalization of the 2-D results to a space of higher dimension. In an N-dimensional space ($N \geq 2$), equation (3) may be replaced by

$$Q = \sum_{\mu,\nu=1}^{N} \alpha_{\mu\nu} \frac{\partial^2}{\partial x_\mu \partial x_\nu} \tag{42}$$

where $\alpha_{\mu\nu}$ are real constants and x_μ the independent variables. Furthermore, the elliptic condition (4) is replaced by the requirement that the matrix

$$A \overset{\text{def.}}{=} (\alpha_{\mu\nu}) \tag{43}$$

is symmetric and positive definite (SPD). Also the operator P assumes the new form

$$P \overset{\text{def.}}{=} \sum_{\mu=1}^{N} p_\mu \frac{\partial^2}{\partial x_\mu \partial x_\mu} \qquad (p_\mu > 0,\ \mu = 1,2,3,\ldots,N) \tag{44}$$

With the aid of equations (42) to (44) and theorem 1 in appendix B, equations (6) to (37) may be generalized in a straightforward manner. However, it should be cautioned that, for $N \geq 2$, the parameters λ_{max} and λ_{min} are defined, respectively, as the greatest and the smallest eigenvalues of the SPD matrix

$$\hat{A} \overset{\text{def.}}{=} (\hat{\alpha}_{\mu\nu}) \tag{45}$$

where

$$\hat{\alpha}_{\mu\nu} \overset{\text{def.}}{=} \frac{\alpha_{\mu\nu}}{\sqrt{p_\mu p_\nu}} \tag{46}$$

Finally, it is noted that equations (38) to (41) have no trivial counterparts in a space with $N > 2$.

Local Relaxation

In this section, the numerical procedure developed in the previous section is extended to solve PDE's with variable coefficients. To proceed, the operator Q is initially assumed to have the form defined in equation (3), with the understanding that the coefficients a, b, and c are functions of x and y subjected to the elliptic condition (4).

In the variable coefficient (VC) version of the iterative procedure (7), the operator $\tilde{Q}$ will be defined by using equation (8), with the understanding that the coefficients a, b, and c, respectively, are replaced by a_{ij}, b_{ij}, and c_{ij}. That is, the discretized values of a, b, and c at the grid point (i,j). On the other hand, the coefficients a_o and c_o associated with the operator $\tilde{P}$ (eq. (9)) are again assumed to be positive constants.

The preceding definitions of $\tilde{P}$ and $\tilde{Q}$ are directly applicable to any internal grid point. On a periodic boundary, they are also applicable if the periodic conditions are invoked. Similarly, by using an extrapolation technique (ref. 12), the operators $\tilde{P}$ and $\tilde{Q}$ can be defined on a Neumann boundary.

The relaxation factor τ, in the VC version, is replaced by its grid-point dependent version τ_{ij}. Ideally, the values of τ_{ij}'s may be chosen such that the parameter M^∞ (eq. (24)) is minimized. Unfortunately, this approach is impractical because of the complexity arising from the variable nature of the coefficients of $\tilde{Q}$ and the necessity to consider the boundary conditions. The alternative adopted in the current study is based on the following heuristic arguments: Recall that the analysis described in the previous section is a rigorous von Neumann analysis for equation (10). The results of this analysis are fully justified only under very restricted conditions. However, it is well known that the von Neumann analysis often gives useful results even when its application cannot be fully justified. Particularly, by freezing the variable coefficients at their values at the grid point under consideration, this analysis has been routinely used in the stability study of the numerical procedure solving PDE's with variable coefficients. Because of the above considerations, the VC version of equation (38) is assumed to be

$$\tau_{i,j} = \frac{2}{\hat{a}_{ij} + \hat{c}_{ij}} \tag{47}$$

where

$$\hat{a}_{ij} \overset{\text{def.}}{=} \frac{a_{ij}}{a_o} > 0 \quad \text{and} \quad \hat{c}_{ij} \overset{\text{def.}}{=} \frac{c_{ij}}{c_o} > 0$$

In view of equations (25) and (36), it is also assumed that

$$\lim_{n \to +\infty} \frac{|e_{i,j}^{n+1}|}{|e_{i,j}^{n}|} \approx G_{ij} \tag{48}$$

where G_{ij} is the local error multiplication factor defined by

$$G_{ij} \overset{\text{def.}}{=} \frac{\Sigma_{ij}^* - 1}{\Sigma_{ij}^* + 1} \tag{49}$$

The parameter Σ_{ij}^* is the grid-point dependent version of Σ^*. It will be evaluated by using equations (37) and (30) to (32) with the understanding that the coefficients a, b, and c, respectively, are replaced by a_{ij}, b_{ij}, and c_{ij}. Let

$$\|e^n\| \overset{\text{def.}}{=} \left[\sum_{(i,j)\in\Phi} \left(e^n_{i,j}\right)^2 \right]^{1/2} \tag{50}$$

and

$$G^\infty \overset{\text{def.}}{=} \operatorname*{Max}_{(i,j)\in\Phi} \{G_{ij}\} \tag{51}$$

where Φ denotes the set of (i,j)'s where $u_{i,j}$'s are to be solved. Then, with the aid of equation (24), assumption (48) implies

$$M^\infty \approx G^\infty \tag{52}$$

Several comments can be made relating to equation (52).

(1) Since G^∞ can be evaluated by using the known coefficients a_o, c_o, a_{ij}, b_{ij}, and c_{ij}, the value of M^∞, and thus the convergence rate of the current iterative procedure, can be predicted by using equation (52).

(2) As long as the coefficients a_{ij}, b_{ij}, and c_{ij} do not vary greatly from one grid point to its neighbors, the value of G^∞ is not sensitive to the grid-cell size and aspect ratio. This observation coupled with equation (52) implies that the convergence behavior of the current numerical procedure, generally, may not be sensitive to the grid-cell size and aspect ratio.

(3) The VC version of equation (7) can be expressed in a form in which the coefficients a_o and c_o appear only in the ratio c_o/a_o. As a result, the convergence behavior of the current iterative procedure is dependent on the ratio c_o/a_o, but not on the individual values of a_o and c_o. Similarly, one can also show that the parameter G^∞ is dependent on the ratio c_o/a_o, but not on the individual values of a_o and c_o. Equation (52) suggests that, in order to maximize the convergence rate, the ratio c_o/a_o should be chosen such that G^∞ is at its minimum.

Evaluation of the optimal value of c_o/a_o, generally, may involve complicated numerical calculations. However, in case that $b_{ij} = 0$ for all $(i,j) \in \Phi$, it is shown in appendix D that G^∞ reaches its minimum

$$G^\infty_{\min} \overset{\text{def.}}{=} \frac{\sqrt{\beta_{\max}/\beta_{\min}} - 1}{\sqrt{\beta_{\max}/\beta_{\min}} + 1} \tag{53}$$

if and only if

$$\frac{c_o}{a_o} = \sqrt{\beta_{\max} \cdot \beta_{\min}} \tag{54}$$

where

$$\left.\begin{aligned} \beta_{\max} &\overset{\text{def.}}{=} \operatorname*{Max}_{(i,j)\in\Phi} \left\{\frac{c_{ij}}{a_{ij}}\right\} \\ \beta_{\min} &\overset{\text{def.}}{=} \operatorname*{Min}_{(i,j)\in\Phi} \left\{\frac{c_{ij}}{a_{ij}}\right\} \end{aligned}\right\} \tag{55}$$

may be evaluated either analytically or numerically.

The current procedure can be modified to solve a class of self-adjoint PDE's. That is,

$$Q = Q^+(x,y) \overset{\text{def.}}{=} \frac{\partial}{\partial x}\left(p(x,y)\frac{\partial}{\partial x}\right) + \frac{\partial}{\partial y}\left(q(x,y)\frac{\partial}{\partial y}\right) \tag{56}$$

where p and q are arbitrary positive functions of x and y. A central difference operator $\tilde{Q}^+$ corresponding to the differential operator Q^+ is defined by (ref. 13)

$$\begin{aligned} \tilde{Q}^+(v_{i,j}) \overset{\text{def.}}{=} (\Delta x)^{-2}\Big[& p_{(i-1/2)j}v_{i-1,j} + p_{(i+1/2)j}v_{i+1,j} \\ & - \left(p_{(i-1/2)j} + p_{(i+1/2)j}\right)v_{i,j}\Big] \\ + (\Delta y)^{-2}\Big[& q_{i(j-1/2)}v_{i,j-1} + q_{i(j+1/2)}v_{i,j+1} \\ & - \left(q_{i(j-1/2)} + q_{i(j+1/2)}\right)v_{i,j}\Big] \end{aligned} \tag{57}$$

where

$$p_{(i\pm1/2)j} \overset{\text{def.}}{=} p(x_i \pm \Delta x/2, y_j)$$

and

$$q_{i(j\pm1/2)} \overset{\text{def.}}{=} q(x_i, y_j \pm \Delta y/2)$$

With the assumption that coefficients p and q do not vary greatly from one grid point to its neighbors, then

$$\begin{aligned} \tilde{Q}^+(v_{i,j}) \doteq & \, p_{ij}(\Delta x)^{-2}(v_{i-1,j} + v_{i+1,j} - 2v_{i,j}) \\ & + q_{ij}(\Delta y)^{-2}(v_{i,j-1} + v_{i,j+1} - 2v_{i,j}) \end{aligned}$$

Thus $\tilde{Q}^+(v_{i,j}) \doteq \tilde{Q}(v_{i,j})$ (eq. (8)), if $a_{ij} = p_{ij}$, $c_{ij} = q_{ij}$, and $b_{ij} = 0$ for all $(i,j) \in \Phi$. This observation coupled with equation (47) lead to the assumption

$$\tau_{ij} = \frac{2}{\hat{p}_{ij} + \hat{q}_{ij}} \tag{58}$$

where $\hat{p}_{ij} \overset{\text{def.}}{=} p_{ij}/a_o$ and $\hat{q}_{ij} \overset{\text{def.}}{=} q_{ij}/c_o$. Similarly, in the case that $Q = Q^+(x,y)$, the parameter G^∞ will be evaluated by

assuming $a_{ij} = p_{ij}$, $c_{ij} = q_{ij}$, and $b_{ij} = 0$. Also the right sides of equations (53) and (54) will be evaluated with

$$\left.\begin{aligned} \beta_{\max} &\overset{\text{def.}}{=} \operatorname*{Max}_{(i,j)\in\Phi}\left\{\frac{q_{ij}}{p_{ij}}\right\} \\ \beta_{\min} &\overset{\text{def.}}{=} \operatorname*{Min}_{(i,j)\in\Phi}\left\{\frac{q_{ij}}{p_{ij}}\right\} \end{aligned}\right\} \tag{59}$$

The technique of local relaxation described for 2-D problems, can be applied in a similar fashion, to 3-D problems. The value of this technique as a tool to solve PDE's with variable coefficients will be demonstrated in the subsequent sections.

Numerical Evaluation

Numerical evaluation of the current method begins with the following preliminaries:

(1) In this section the domain for all numerical problems is assumed to be $1 \geq x \geq 0$ and $1 \geq y \geq 0$. Moreover, the operator $\widetilde{P}$ is inverted by using a Fast Poisson Solver (ref. 14).

(2) The convergence rate is evaluated by using

$$O_e(n) \overset{\text{def.}}{=} -\log_{10}\left(\frac{\|e^n\|}{\|e^0\|}\right) \tag{60}$$

or

$$O_r(n) \overset{\text{def.}}{=} -\log_{10}\left(\frac{\|r^n\|}{\|r^0\|}\right) \tag{61}$$

where e^n is the error norm defined in equation (50), while $\|r^n\|$ is the residual norm defined by

$$\|r^n\| \overset{\text{def.}}{=} \left\{ \sum_{(i,j)\in\Phi} \left[\widetilde{Q}\left(u^n_{i,j}\right) - h_{i,j}\right]^2 \right\}^{1/2} \tag{62}$$

The solution $u_{i,j}$ obtained to machine accuracy is used to evaluate $O_e(n)$. Furthermore, since $u^0_{i,j} = 0$ for all $(i,j) \in \Phi$ in the current numerical study, $O_e(n)$ can be interpreted as the number of correct digits in $u^n_{i,j}$.

(3) In view of equation (52), the parameters $O_e(n)$ and $O_r(n)$ will be predicted by using

$$O_t(n) = -n \cdot \log_{10}(G^\infty) \tag{63}$$

or its continuous version; that is,

$$O^*_t(n) \overset{\text{def.}}{=} \lim_{\Delta x, \Delta y \to 0} O_t(n) \tag{64}$$

Numerical evaluation involving PDE's with constant coefficients are given in reference 15. The first group of PDE's to be studied includes

$$(1 + 2x^2 + 2y^2)\frac{\partial^2 u}{\partial x^2} + (1 + x^2 + y^2)\frac{\partial^2 u}{\partial x^2} = 1 \tag{65}$$

$$(1 + 2x^2 + 2y^2)\frac{\partial^2 u}{\partial x^2} + (1 + x^2 + y^2)\frac{\partial^2 u}{\partial x \partial y} + (1 + x^2 + y^2)\frac{\partial^2 u}{\partial y^2} = 1 \tag{66}$$

$$(1 + 2x^2 + 2y^2)\frac{\partial^2 u}{\partial x^2} + (1 + x^2 + y^2)\frac{\partial^2 u}{\partial x \partial y} + (1 + x^2 + y^2)\frac{\partial^2 u}{\partial y^2} + \left[1 + 3e^{(x^2+y^2)}\right]\frac{\partial u}{\partial x} + \left[1 + 3e^{(x^2+y^2)}\right]\frac{\partial u}{\partial y} - \left[1 + 3e^{(x^2+y^2)}\right]u = 1 \tag{67}$$

$$\frac{\partial}{\partial x}\left[\left\{1 + (x^2 + y^2)^\ell\right\}\frac{\partial u}{\partial x}\right] + \frac{\partial}{\partial y}\left[\left\{2 + (x^2 + y^2)^\ell\right\}\frac{\partial u}{\partial y}\right] = 1 \qquad \ell = 2,4,6,8 \tag{68}$$

Fifteen numerical problems associated with the above PDE's are defined in table I. The parameters *MX* and *MY*, respectively, are the numbers of grid intervals in the x and y directions. The other parameter *IB* specifies the particular set of boundary conditions (fig. 1). All these problems are solved by assuming $a_o = c_o = 1$.

Problems 1 to 5 are associated with the same PDE (65). They differ on the grid-cell size, aspect ratio, and boundary conditions. As shown in table I, the values of either $O_t(20)$ or $O^*_t(20)$ are fairly accurate estimates of $O_r(20)$. Also, as expected from the current theoretical development and the experiences of other researchers (refs. 7 and 8), the effects of grid-cell size and aspect ratio on the convergence rate are minimal. Even the very large aspect ratio (16:1) does not cause any significant reduction in the convergence rate. Furthermore, the convergence rate is insensitive to the particular set of boundary conditions used.

TABLE I.—NUMERICAL PROBLEMS ASSOCIATED WITH EQUATIONS (65) TO (68) AND COMPARISONS OF $O_r(n)$, $O_t(n)$, AND $O_t^*(n)$

[$n = 20$ for problems 1 to 5, $n = 32$ for problems 6 to 15.]

Problem number	Equation	ℓ	IB	MX	MY	$O_r(n)$	$O_t(n)$	$O_t^*(n)$
1	(65)	NA	1	16	16	14.50	12.34	12.04
2	↓	NA	1	64	64	13.78	12.11	12.04
3		NA	1	64	4	13.96	12.68	12.04
4		NA	2	16	16	13.96	12.34	12.04
5	↓	NA	3	16	16	13.62	12.18	12.04
6	(66)	NA	1	16	16	14.43	9.69	9.63
7	↓	NA	1	64	64	12.58	9.64	9.63
8		NA	1	64	4	19.04	10.01	9.63
9		NA	2	16	16	14.50	9.66	9.63
10	↓	NA	3	16	16	15.44	9.66	9.63
11	(67)	NA	1	16	16	13.05	9.69	9.63
12	(68)	2	1	16	16	17.24	15.27	15.27
13	↓	4	1	16	16	16.83	15.27	15.27
14		6	1	16	16	13.44	15.27	15.27
15	↓	8	1	16	16	7.55	15.27	15.27

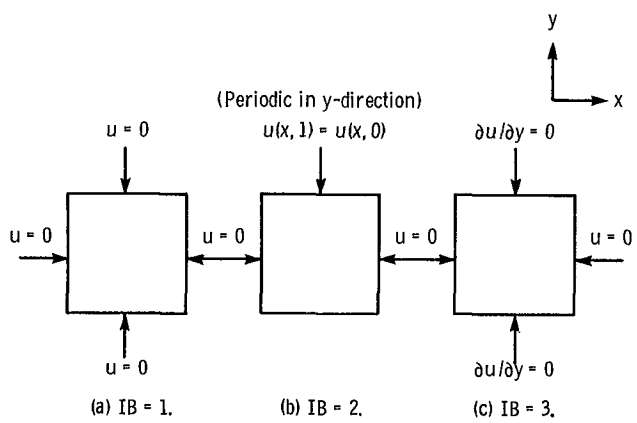

Figure 1.—Three sets of boundary conditions on a unit square.

Problems 6 to 10 are associated with equation (66), which differs from equation (65) only in the appearance of a cross-derivative term. The numerical results indicate that the convergence rate may be substantially underestimated by the parameter $O_t(32)$ or $O_t^*(32)$. Furthermore, it is more sensitive to the change of grid-cell size and aspect ratio. An explanation for these peculiar behaviors associated with a PDE with a cross-derivative term is given at the end of appendix B.

The success of the current numerical method in solving a PDE with a cross derivative term is rather significant. This author is unaware of any earlier work which solves PDE's of this type with a semidirect procedure. The lack of progress in this area may be due to the fact that it is very difficult to choose a separable operator P which closely resembles a nonseparable operator Q containing a cross-derivative term. (By definition, a separable operator P can not have a cross-derivative term.)

Equation (67) contains first-order and zero-order derivative terms. This type of PDE is solved by simply adding the central difference form of those terms to the term $\tilde{Q}(u_{i,j}^n)$ in equation (7). The value of $O_r(32)$ for problem 11 indicates that the current procedure works very well even though the coefficients of first-order and zero-order derivative terms in equation (67) are of the same order of magnitude as the second-order terms. This is rather unexpected because the coefficients of lower order terms are completely neglected in the evaluation of the local relaxation factor.

Equation (68) belongs to the class of self-adjoint PDE's defined in equation (56). The variation of the values of the coefficients p and q increases progressively as one goes from $\ell = 2$ to $\ell = 4$ and so on. For $\ell = 8$, the increase in the values of p and q from one corner ($x = y = 0$) to another corner ($x = y = 1$) on the unit square is of the order of 100 times. It might appear that the technique of local relaxation is no longer valid. The results shown in table I indicate the current method is still useful in this extreme case.

The numerical study of problems 1 to 15 concludes with a discussion on their convergence histories. Since equation (63) represents a linear relation between $O_t(n)$ and n, it is not surprising that the relations between $O_r(n)$ and n curves are closely approximated by straight lines for the above problems with the exception of perhaps problems 13 to 15. As shown in figure 2, the linear relation between $O_r(n)$ and n gradually deteriorates as the variation of the coefficients p and q increases. The robustness of the current algorithm is most evident in its ability to reverse the trend toward divergence during the first few iterations.

The second group of PDE's to be studied includes

$$\frac{\partial}{\partial x}\left\{\left[1 + (x+y)^2\right]^2 \frac{\partial u}{\partial x}\right\} + \frac{\partial}{\partial y}\left\{\left[1 + \sin^2(x+y)\right]^2 \frac{\partial u}{\partial y}\right\} = h_1(x,y) \tag{69}$$

$$\frac{\partial}{\partial x}\left\{\left[1 + \frac{1}{2}(x^4 + y^4)\right]^2 \frac{\partial u}{\partial x}\right\} + \frac{\partial}{\partial y}\left\{\left[1 + \frac{1}{2}(x^4 + y^4)\right]^2 \frac{\partial u}{\partial y}\right\} = h_2(x,y) \tag{70}$$

where $h_1(x,y)$ and $h_2(x,y)$ are source terms chosen such that

$$u = u_1(x,y) \overset{\text{def.}}{=} \sin x \sin y \tag{71}$$

and

$$u = u_2(x,y) \overset{\text{def.}}{=} [x(1-x)y(1-y)]^2 \tag{72}$$

respectively, are the exact solutions of equations (69) and (70).

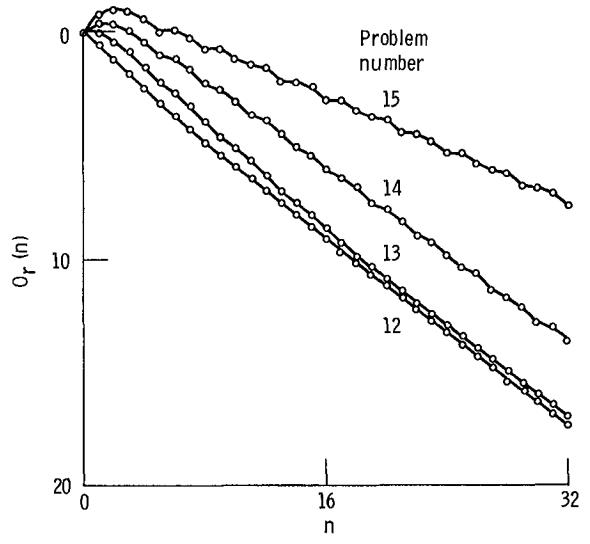

Figure 2.—Convergence histories of problem numbers 12 to 15.

Definitions of the four numerical problems associated with equations (69) and (70) along with the values of $O_e(10)$, $O_r(10)$, and $O_t(10)$ are given in table II. The boundary values of u in these problems are specified by using equation (71) or equation (72).

Problem 16 is one of the test problems used by Bank (ref. 8). Compared with the current value of $O_e(10) = 5.96$, the values obtained by Bank are 3.49 without using any scaling technique, and 3.87, 4.81, and 6.76 using three different scaling functions. Since the operator $\tilde{P}$ used in Bank's method is a general separable operator, the corresponding FDS code usually must be made to individual specifications and is about five times slower than that for the Laplacian ∇^2. Thus, the current algorithm is easier to use and, for problem 16, more efficient by at least a factor of 4.

Problem 18 is another test problem used by Bank. Compared with the current value of $O_e(10) = 8.59$, the value obtained by Bank is 5.88 without using his scaling technique and 14.79 if a scaling function is used. This problem along with problem 19 was also solved by Concus and Golub (ref. 7). The method of Concus and Golub is also driven by a fast Poisson solver and the results obtained are comparable with ours. However, their method is applicable only when $p = q$ as in the case of equation (70).

TABLE II.—NUMERICAL PROBLEMS ASSOCIATED WITH EQUATIONS (69) and (70), AND COMPARISONS OF $O_e(10)$, $O_r(10)$, and $O_t(10)$

Problem number	Equation	MX	MY	c_o/a_o	$O_e(10)$	$O_r(10)$	$O_t(10)$
16	(69)	16	16	[a]0.423	5.96	5.33	3.92
17	(69)	64	64	[a]0.379	5.27	4.22	3.47
18	(70)	16	16	[a]1.0	8.59	8.09	∞
19	(70)	64	64	[a]1.0	8.47	7.95	∞

[a]Evaluated from equation (54).

The last PDE's to be studied in this section are

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial}{\partial y}\left\{\left[1 + \frac{1}{2}(x - y)\right]\frac{\partial u}{\partial y}\right\} = 0 \tag{73}$$

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial}{\partial y}\left\{\left[10 + \frac{1}{2}(x - y)\right]\frac{\partial u}{\partial y}\right\} = 0 \tag{74}$$

An exact solution for both equations (73) and (74) is

$$u = y + \frac{1}{4}x^2 \tag{75}$$

Equations (73) and (74) were numerically solved by using a grid with $MX = MY = 31$. It is assumed that the values of u at all boundary grid points are specified by using equation (75). To compare the efficiency of the current method with traditional iterative methods, these numerical problems are solved by using the current method along with SLOR (successive line overrelaxation). The results of central processing unit (CPU) time comparisons are summarized in table III. Those parameters used in this table which were not defined previously are as follows:

N_c smallest value of n which satisfies the convergence criterion

$$(\Delta x)^2 \cdot \|r^n\| < 10^{-8} \tag{76}$$

where $\|r^n\|$ is defined in equation (62) and $\Delta x = 1/31$

T_f CPU time used in the execution of the FDS code

T_t total CPU time (IBM 370/3033AP) used to satisfy the convergence criterion (76)

ω_o optimal value of the relaxation factor used in the SLOR method (determined by repeated numerical experiments)

According to table III, the total CPU time required for the solution of either equation (73) or equation (74) with SLOR is about twice that with the current method. This comparison becomes even more favorable toward the current method if

TABLE III.—CPU TIME COMPARISONS BETWEEN CURRENT METHOD AND SLOR METHOD

Equation	Solution method	c_o/a_o	ω_o	N_c	T_t, sec	T_f, sec
(73)	Current	[a]0.8839	NA	13	1.871	1.345
(73)	SLOR	NA	1.752	83	3.790	NA
(74)	Current	[a]9.989	NA	6	0.888	0.614
(74)	SLOR	NA	1.510	44	2.039	NA

[a]Evaluate from equation (54).

one recalls that the prediction of ω_o is elusive. A small error in this prediction may result in a large increase in the value of T_t. For example, in the solution of equation (73) with SLOR, a change of the value of the relaxation factor from 1.752 to 1.680 results in an increase in the value of T_t from 3.790 to 6.565 seconds. On the other hand, as shown in table IV, the optimal value of c_o/a_o evaluated by using equation (54) usually is very accurate. Moreover, since the fast Poisson solver (ref. 14) currently used is a general purpose code, the value of T_f can be reduced further if the fast Poisson solver is optimized.

To conclude this section, the current local relaxation procedure is compared numerically with a procedure which differs from the former only in the use of a constant relaxation factor τ_c. With the assumption of $a_o = c_o = 1$, problem 16 was solved with different values of τ_c. As shown in table V, $O_e(10)$ reaches its best value ($\doteqdot 2.278$) at $\tau_c \doteqdot 0.103$. Even this best value is substantially below that ($\doteqdot 3.47$) obtained by using the local relaxation method ($a_o = c_o = 1$). Furthermore, the accurate prediction of optimal τ_c is by no means easy (e.g., pp. 964 to 966 of ref. 8). Thus the current procedure has a clear edge over a procedure that uses a constant relaxation factor.

Application to a Three-Dimensional Flow Problem

In this section, the current semidirect procedure is incorporated into an Euler solver (ref. 9) to obtain the inviscid solution for 3-D steady incompressible rotational flow in a 180° turning channel (fig. 3).

The Euler solver is formed by the inner and the outer loops. The inner loop solves the elliptic equations, while the outer loop solves the hyperbolic equations. In each pass through the inner loop, the velocity field $\vec{V}$ is updated to satisfy the continuity equation

$$\vec{\nabla} \cdot \vec{V} = 0 \tag{77}$$

and the velocity-vorticity relation

$$\vec{\nabla} \times \vec{V} = \vec{\Omega} \tag{78}$$

where $\vec{\Omega}$ is a known divergence-free (i.e., $\vec{\nabla} \cdot \vec{\Omega} = 0$) vorticity field. In the current study, a solution procedure different from that described in reference 9 is used to solve equations (77) and (78). The general solution of equations (77) and (78) can be expressed as

$$\vec{V} = \vec{V}_s + \vec{\nabla} u \tag{79}$$

where $\vec{V}_s$ is any special solution of equation (78) and u is a solution of

$$\nabla^2 u = -\vec{\nabla} \cdot \vec{V}_s \tag{80}$$

As a result, once a special solution $\vec{V}_s$ is obtained (ref. 9), the solution of equation (80) becomes the focal point of the inner-loop calculations.

Let the coordinates $(\bar{x}_1,\bar{x}_2,\bar{x}_3)$ refer to physical space and (x_1,x_2,x_3) to computational space. It is shown in reference 9 that the turning channel in figure 3 is a mapping of a parallelepiped ($2 \geq x_1 \geq -2$, $0.75 \geq x_2 \geq 0.65$, $0.1 \geq x_3 \geq 0$) in computational space. In physical space, equation (80) is a Poisson's equation. However, it cannot be solved by using the current procedure, since the physical domain is not a parallelepiped. On the other hand, in computational space, the domain is a parallelepiped and equation (80) becomes

$$\frac{\partial^2 u}{\partial x_1^2} + \frac{\partial^2 u}{\partial x_2^2} + \eta \frac{\partial^2 u}{\partial x_3^2} = -\left(\frac{\partial V_{s,1}}{\partial x_1} + \frac{\partial V_{s,2}}{\partial x_2} + \eta \frac{\partial V_{s,3}}{\partial x_3} \right) \tag{81}$$

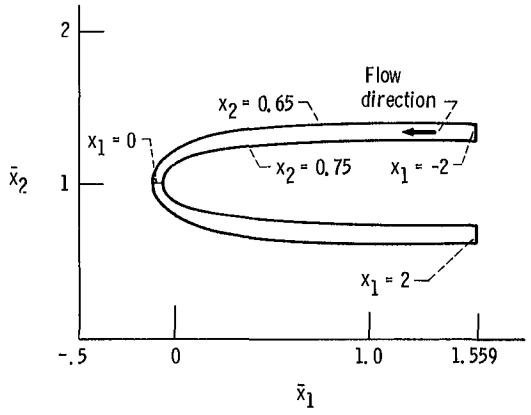

Figure 3.—A converging and diverging turning channel ($\bar{x}_3$ is suppressed).

TABLE IV.—N_c AND $O_r(13)$ AS FUNCTIONS OF c_o/a_o IN THE NUMERICAL SOLUTION OF EQUATION (73)

c_o/a_o	N_c	$O_r(13)$
0.8	14	8.3343
[a].8839	13	9.4040
.895	13	9.4891
.897	13	9.4955
.899	13	9.4987
[b].900	13	9.4990
.901	13	9.4985
.903	13	9.4949
.905	13	9.4882
1.0	14	8.4831

[a]Evaluated from equation (54).
[b]Actual optimal value.

TABLE V.—$O_e(10)$ AS A FUNCTION OF τ_c FOR PROBLEM 16

τ_c	$O_e(10)$
0.02	0.6
.06	1.514
.10	2.259
.102	2.276
.103	2.278
.104	2.272
.105	2.255
.11	1.989
.12	1.2
.13	(a)

[a]Do not converge if $\tau_c \geq 0.13$.

where $V_{s,1}, V_{s,2}$, and $V_{s,3}$ are the covariant components of the known vector field $\vec{V}_s$ and

$$\eta = \eta(x,y) \overset{\text{def.}}{=} \frac{\cosh(\pi x_1) + \cos(\pi x_2)}{\cosh(\pi x_1) - \cos(\pi x_2)} \tag{82}$$

The 3-D operator Q associated with equation (81) is a special case of that defined in equation (42). Let the parameter G^∞ and the set Φ for 3-D problems be defined similar to their counterparts for 2-D problems. Using a line of arguments similar to that presented in appendix D shows that the parameter G^∞ reaches its minimum

$$G^\infty_{\min} \overset{\text{def.}}{=} \frac{\sqrt{\eta_{\max}/\eta_{\min}} - 1}{\sqrt{\eta_{\max}/\eta_{\min}} + 1} \tag{83}$$

if and only if the coefficients of the operator P (eq. (44)) are chosen such that

$$\frac{p_3}{p_1} = \frac{p_3}{p_2} = \sqrt{\eta_{\max} \cdot \eta_{\min}} \tag{84}$$

Here $\eta_{\max}$ and $\eta_{\min}$, respectively, are the maximum and minimum of the function η in Φ.

In a successful effort to obtain a secondary flow solution in the turning channel, equation (81) was solved once during each of 25 passes through the inner loop (The source terms on the right side of equation (81) vary from one pass to another.) The central difference form of equation (81) is obtained by using a grid with 144 uniform intervals in the x_1-direction and 12 uniform intervals in both the x_2- and x_3-directions. It is assumed that the normal derivative of u vanishes at all boundaries except at the exit plane ($x_1 = 2$) where $u = 0$. Thus $\eta_{\max} = \eta(-2,0.65) \doteqdot 0.9966$ and $\eta_{\min} = \eta(0,0.75) \doteqdot 0.1716$. As suggested by equation (84), equation (81) was solved by assuming $P_1 = P_2 = 1$ and $P = 0.4135$. The values of $O_r(8)$ obtained range from 3.69 to 4.48. All are higher than the value of $O_t(8) \doteqdot 3.07$ as evaluated from equation (83). The convergence rates achieved by using the current procedure are dramatic improvements over those obtained in reference 9.

Concluding Remarks

An efficient semidirect procedure for solving a large class of nonseparable elliptic problems has been developed. In applying this method, a user simply evaluates the terms on the right side of equation (7) and uses them as the input for a fast Poisson solver. The user is not required to deal with a large sparse matrix as in the case of a traditional iterative procedure.

The local relaxation factor is evaluated by using an algebraic formula. This formula along with a convergence rate prediction method is developed by using a heuristic argument. It is shown numerically that the prediction method is an effective tool for estimating the convergence rate.

The convergence rate can be accelerated by optimizing the coefficients of the finite-difference operator $\tilde{P}$. It is shown that this optimization can be carried out easily for a large class of elliptic PDE's.

It is also shown that the convergence rate of the current procedure is relatively insensitive to the grid-cell size and aspect ratio. The underlying reason for this insensitivity is the existence of the uniform bounds $\lambda_{\max}$ and $\lambda_{\min}$. Their existence also contributes greatly to the simplicity of the current procedure.

Although not shown in this report, the current procedure may also be used to solve a second order quasi-linear elliptic PDE. Since the coefficients of this PDE are functions of the dependent variable and its derivatives, the local relaxation factor must be updated during each iteration for this type of application.

National Aeronautics and Space Administration
Lewis Research Center
Cleveland, Ohio, January 13, 1986

Appendix A
Derivation of Equation (21)

In this appendix, we will derive equation (21) by using equations (10) to (20). To proceed, we define the sets $(n = 0,1,2,\ldots)$

$$e^n \overset{\text{def.}}{=} \left\{e_{ij}^n \middle| i = 0,1,2,\ldots,(K-1);\ j = 0,1,2,\ldots,(L-1)\right\} \tag{A1}$$

As a result of equations (12) and (13), for a given $n \geq 0$, any $e_{i,j}^n$ $(i,j = 0, \pm 1, \pm 2,\ldots)$ is equal to an element of e^n. Each e^n contains $K \times L$ elements and, therefore, can be considered as a vector in a $(K \times L)$-dimensional vector space $C^{(K\times L)}$. Similarly, since

$$\varphi_{i,j}^{(k,\ell)} = \varphi_{i+K,j}^{(k,\ell)} = \varphi_{i,j+L}^{(k,\ell)} \quad (i,j = 0, \pm 1, \pm 2,\ldots) \tag{A2}$$

any $\varphi_{i,j}^{(k,\ell)}$ $(i,j = 0, \pm 1, \pm 2,\ldots)$ is equal to an element of the set

$$\varphi^{(k,\ell)} \overset{\text{def.}}{=} \left\{\varphi_{i,j}^{(k,\ell)} \middle| i = 0,1,2,\ldots,(K-1);\right.$$
$$\left. j = 0,1,2,\ldots,(L-1)\right\} \tag{A3}$$

Each $\varphi^{(k,\ell)}$ can also be considered as a vector in $C^{(K\times L)}$. It can be shown that the set of $(K \times L)$ vectors

$$\left\{\varphi^{(k,\ell)} \middle| k = 0,1,2,\ldots,(K-1);\ \ell = 0,1,2,\ldots,(L-1)\right\} \tag{A4}$$

forms an orthonormal basis for $C^{(K\times L)}$; that is,

$$\sum_{i=0}^{(K-1)} \sum_{j=0}^{(L-1)} \varphi_{i,j}^{(k,\ell)} \overline{\varphi_{i,j}^{(k',\ell')}} = \delta_{kk'}\delta_{\ell\ell'}$$

$$(k,k' = 0,1,2,\ldots,(K-1);\ \ell,\ell' = 0,1,2,\ldots,(L-1)) \tag{A5}$$

where $\delta_{kk'}$ is the Kronecker delta symbol. Also, it can be shown that

$$\widetilde{P}\left(\varphi_{i,j}^{(k,\ell)}\right) = \sigma_p^{(k,\ell)}\varphi_{i,j}^{(k,\ell)} \tag{A6}$$

and

$$\widetilde{Q}\left(\varphi_{i,j}^{(k,\ell)}\right) = \sigma_q^{(k,\ell)}\varphi_{i,j}^{(k,\ell)} \tag{A7}$$

where $\widetilde{P}$, $\widetilde{Q}$, $\sigma_p^{(k,\ell)}$, and $\sigma_q^{(k,\ell)}$ are defined in the section Analysis.

Each vector e^n can be expressed as a linear combination of $\varphi^{(k,\ell)}$'s. This fact coupled with equations (12), (13), (A2), and (A5) implies that

$$e_{i,j}^n = \sum_{k=0}^{(K-1)} \sum_{\ell=0}^{(L-1)} E^{n,(k,\ell)}\varphi_{i,j}^{(k,\ell)}$$

$$(n = 0,1,2,\ldots;\ i,j = 0, \pm 1, \pm 2,\ldots) \tag{A8}$$

where

$$E^{n,(k,\ell)} \overset{\text{def.}}{=} \sum_{i=0}^{(K-1)} \sum_{j=0}^{(L-1)} e_{i,j}^n \overline{\varphi_{i,j}^{(k,\ell)}} \tag{A9}$$

Substituting equation (A8) into (10) and using equations (A5) to (A7), one obtains

$$\left(E^{n+1,(k,\ell)} - E^{n,(k,\ell)}\right)\sigma_p^{(k,\ell)} = -\tau E^{n,(k,\ell)}\sigma_q^{(k,\ell)}$$

$$(n = 0,1,2,\ldots;\ k = 0,1,2,\ldots,(K-1);$$
$$\ell = 0,1,2,\ldots,(L-1)) \tag{A10}$$

For $k = \ell = 0$, equation (A10) is an identity since $\sigma_p^{(0,0)} = \sigma_q^{(0,0)} = 0$. On the other hand, since $\sigma_p^{(k,\ell)} < 0$ if $(k,\ell) \in \Psi$, equation (A10) implies that

$$E^{n+1,(k,\ell)} = \left[G^{(k,\ell)}(\tau)\right] \cdot E^{n,(k,\ell)} \qquad (k,\ell) \in \Psi \tag{A11}$$

where $G^{(k,\ell)}(\tau)$ and Ψ, respectively, are defined in equations (19) and (20). Equation (21) is an immediate result of equations (A8) and (A11) and the assumption

$$E^{n,(0,0)} = 0 \qquad (n = 1,2,3,\ldots) \tag{A12}$$

It should be noted that equation (A12) is introduced to ensure the uniqueness of the solution. Using equation (A9) and the fact that $\varphi_{i,j}^{(0,0)} = 1/\sqrt{KL}$ for $i,j = 0, \pm 1, \pm 2,\ldots$, it can be shown that equation (A12) is equivalent to equation (14).

Appendix B
Proof of Expressions (33) and (34)

Expressions (33) and (34) will be established by using the following theorems:

Theorem 1

Let λ_1 and λ_N, respectively, be the greatest and the smallest eigenvalues of an $N \times N$ real symmetric matrix $D \stackrel{\text{def.}}{=} (d_{\mu\nu})$. Let real vectors $s = (s_1,s_2,\ldots,s_N)$ and $t = (t_1,t_2,\ldots,t_N)$ satisfy

$$\sum_{\mu=1}^{N} (s_\mu)^2 = 1 \tag{B1}$$

and

$$(t_\mu)^2 \leq 1 \qquad (\mu = 1,2,\ldots,N) \tag{B2}$$

Then the following may be stated:

$$(1)\ \lambda_1 \geq (D;s,t) \stackrel{\text{def.}}{=} \sum_{\mu=1}^{N} d_{\mu\mu}(s_\mu)^2 + \sum_{\mu=1}^{N} \sum_{\substack{\nu=1 \\ \nu\neq\mu}}^{N} d_{\mu\nu}s_\mu s_\nu t_\mu t_\nu \geq \lambda_N \tag{B3}$$

(2) If, in addition, $\lambda_1 \neq d_{\mu\mu}$ for $\mu = 1,2,\ldots,N$, then $(D;s,t) = \lambda_1$ if and only if $s_\mu\sqrt{1-(t_\mu)^2} = 0$, for $\mu = 1,2,\ldots,N$, and $(s_1t_1,s_2t_2,\ldots,s_Nt_N)$ is an eigenvector of the matrix D with eigenvalue λ_1. This statement remains true if λ_1 is replaced by λ_N.

(3) If δ $(1 \leq \delta \leq N)$ is an integer such that $d_{\delta\delta} = \lambda_1$, then $d_{\delta\nu} = 0$ for all $\nu \neq \delta (1 \leq \nu \leq N)$. As a result, $(D;s,t)$ is independent of the component t_δ. This statement remains true if λ_1 is replaced by λ_N.

This theorem is a special case of theorem 1 in reference 16.

Theorem 2

Let $\hat{a}$, $\hat{b}$, and $\hat{c}$ be the constants defined in expression (32). Let

$$F(s_1,s_2,t_1,t_2) \stackrel{\text{def.}}{=} \hat{a}(s_1)^2 + \hat{c}(s_2)^2 + 2\hat{b}s_1s_2t_1t_2 \tag{B4}$$

where s_1, s_2, t_1, and t_2 are real variables and satisfy

$$(s_1)^2 + (s_2)^2 = 1 \tag{B5}$$

$$\left.\begin{aligned} s_1 &\geq 0 \\ s_2 &\geq 0 \end{aligned}\right\} \tag{B6}$$

and

$$\left.\begin{aligned} (t_1)^2 &\leq 1 \\ (t_2)^2 &\leq 1 \end{aligned}\right\} \tag{B7}$$

Then the following may be stated:

(1) $\lambda_{\max} \geq F \geq \lambda_{\min} > 0$ (B8)

where $\lambda_{\max}$ and $\lambda_{\min}$, respectively, are defined in equations (30) and (31).

(2) If $\hat{b} = 0$ and $\hat{a} = \hat{c}$, then $f = \lambda_{\max} = \lambda_{\min}$

(3) If $\hat{b} = 0$ and $\hat{c} > \hat{a}(\hat{a} > \hat{c})$, then
  (a) $F = \lambda_{\max}$ if and only if $s_1 = 0$ $(s_2 = 0)$
  (b) $F = \lambda_{\min}$ if and only if $s_2 = 0$ $(s_1 = 0)$

(4) If $\hat{b} \neq 0$, then
  (a) $F = \lambda_{\max}$ if and only if either

$$s_1 = s_1^+$$

$$s_2 = s_2^+$$

$$t_1 = 1$$

$$t_2 = \frac{\hat{b}}{|\hat{b}|}$$

or

$$s_1 = s_1^+$$

$$s_2 = s_2^+$$

$$t_1 = -1$$

$$t_2 = -\frac{\hat{b}}{|\hat{b}|}$$

where

$$s_1^+ \stackrel{\text{def.}}{=} \frac{|\hat{b}|}{\sqrt{(\lambda_{\max} - \hat{a})^2 + (\hat{b})^2}} > 0 \tag{B9}$$

and

$$s_2^{+\text{def.}} \frac{\lambda_{\max} - \hat{a}}{\sqrt{(\lambda_{\max} - \hat{a})^2 + (\hat{b})^2}} > 0 \qquad \text{(B10)}$$

(b) $F = \lambda_{\min}$ if and only if either

$$s_1 = s_1^-$$

$$s_2 = s_2^-$$

$$t_1 = 1$$

$$t_2 = -\frac{\hat{b}}{|\hat{b}|}$$

or

$$s_1 = s_1^-$$

$$s_2 = s_2^-$$

$$t_1 = -1$$

$$t_2 = -\frac{\hat{b}}{|\hat{b}|}$$

where

$$s_1^{-\text{def.}} \frac{|\hat{b}|}{\sqrt{(\lambda_{\min} - \hat{a})^2 + (\hat{b})^2}} > 0 \qquad \text{(B11)}$$

and

$$s_2^{-\text{def.}} \frac{\hat{a} - \lambda_{\min}}{\sqrt{(\lambda_{\min} - \hat{a})^2 + (\hat{b})^2}} > 0 \qquad \text{(B12)}$$

Proof

Since (1) $\lambda_{\max}$ and $\lambda_{\min}$ are eigenvalues of the matrix

$$\hat{D}^{\text{def.}} \begin{pmatrix} \hat{a} & \hat{b} \\ \hat{b} & \hat{c} \end{pmatrix}$$

and (2) $\lambda_{\max} \geq \lambda_{\min} > 0$, theorem 2 follows directly from theorem 1. QED

To prove expressions (33) and (34), one notes that

$$\gamma^{(k,\ell)\text{def.}} \frac{\sigma_q^{(k,\ell)}}{\sigma_p^{(k,\ell)}} = \hat{a}(s_x)^2 + \hat{c}(s_y)^2 + 2\hat{b}s_x s_y t_x t_y \qquad (k,\ell) \in \Psi \qquad \text{(B13)}$$

where $\sigma_q^{(k,\ell)}$, $\sigma_p^{(k,\ell)}$, $\hat{a}$, $\hat{b}$, $\hat{c}$, and Ψ are defined in the section Analysis, and

$$s_x^{\text{def.}} \frac{\frac{\sqrt{a_o}}{\Delta x} \sin(\pi k/K)}{\sqrt{\left[\frac{\sqrt{a_o}}{\Delta x} \sin\left(\frac{\pi k}{K}\right)\right]^2 + \left[\frac{\sqrt{c_o}}{\Delta y} \sin\left(\frac{\pi \ell}{L}\right)\right]^2}} \qquad (k,\ell) \in \Psi \qquad \text{(B14)}$$

$$s_y^{\text{def.}} \frac{\frac{\sqrt{c_o}}{\Delta y} \sin(\pi \ell/L)}{\sqrt{\left[\frac{\sqrt{a_o}}{\Delta x} \sin\left(\frac{\pi k}{K}\right)\right]^2 + \left[\frac{\sqrt{c_o}}{\Delta y} \sin\left(\frac{\pi \ell}{L}\right)\right]^2}} \qquad (k,\ell) \in \Psi \qquad \text{(B15)}$$

$$t_x^{\text{def.}} \cos(\pi k/K) \qquad (k = 0,1,2,\ldots,(K-1)) \qquad \text{(B16)}$$

and

$$t_y^{\text{def.}} \cos(\pi \ell/L) \qquad (\ell = 0,1,2,\ldots,(L-1)) \qquad \text{(B17)}$$

Using equations (B14) to (B17), one concludes that (1) $(s_x)^2 + (s_y)^2 = 1$, (2) $s_x \geq 0$, $s_y \geq 0$, and (3) $(t_x)^2 \leq 1$, $(t_y)^2 \leq 1$. As a result, part (1) of theorem 2 implies that

$$\lambda_{\max} \geq \gamma^{(k,\ell)} \geq \lambda_{\min} > 0 \qquad (k,\ell) \in \Psi \qquad \text{(B18)}$$

Inequality (33) follows immediately from inequality (B18).

Using parts (1) to (3) of theorem 2 and the facts that (1) $s_x = 0$ if $k = 0$ and $\ell = 1,2,\ldots(L-1)$ and (2) $s_y = 0$ if $\ell = 0$ and $k = 1,2,\ldots(K-1)$, it can be shown that

$$\left.\begin{aligned} \lambda_{\max} &= \gamma_{\max} \qquad (\hat{b} = 0) \\ \lambda_{\min} &= \gamma_{\min} \qquad (\hat{b} = 0) \end{aligned}\right\} \qquad \text{(B19)}$$

Thus, for $\hat{b} = 0$, equation (34) is true and it represents a condition weaker than equation (B19).

To prove equation (34) for $\hat{b} \neq 0$, note that

$$\gamma_1 \stackrel{\text{def.}}{=} \left(s_1^+,\ s_2^+,\ 1,\ \frac{\hat{b}}{|\hat{b}|}\right)$$

$$\gamma_2 \stackrel{\text{def.}}{=} \left(s_1^+,\ s_2^+,\ -1,\ -\frac{\hat{b}}{|\hat{b}|}\right)$$

$$\gamma_3 \stackrel{\text{def.}}{=} \left(s_1^-,\ s_2^-,\ 1,\ -\frac{\hat{b}}{|\hat{b}|}\right)$$

$$\gamma_4 \stackrel{\text{def.}}{=} \left(s_1^-,\ s_2^-,\ -1,\ \frac{\hat{b}}{|\hat{b}|}\right)$$

belong to

$$D(F) \stackrel{\text{def.}}{=} \{(s_1,s_2,t_1,t_2) \mid (s_1)^2 + (s_2)^2 = 1,$$
$$s_1 \geq 0,\ s_2 \geq 0,\ (t_1)^2 \leq 1,\ (t_2)^2 \leq 1\}$$

However, since $(t_1)^2 + (t_2)^2 = 2$ for γ_1, γ_2, γ_3, and γ_4 while $(t_x)^2 + (t_y)^2 < 2$ for all $(k,\ell) \in \Psi$, it follows that γ_1, γ_2, γ_3, and γ_4 do not belong to the set

$$\Gamma \stackrel{\text{def.}}{=} \{(s_x,s_y,t_x,t_y) \mid (k,\ell) \in \Psi\}$$

Thus inequality (B18) and part (4) of theorem 2 imply that

$$\lambda_{\max} > \gamma_{\max} \geq \gamma_{\min} > \lambda_{\min} \qquad (\hat{b} \neq 0) \tag{B20}$$

Furthermore, since (1) the function F is a continuous function over its domain $D(F)$, (2) Γ is a subset of $D(F)$, and (3) for any neighborhood of any one of $\gamma_1,\gamma_2,\gamma_3$, and γ_4, one can find a pair of integers K_o and L_o (see below) such that the intersection of this neighborhood and Γ is not null if $K \geq K_o$ and $L \geq L_o$. (Recall that both Ψ and Γ are dependent on the integers K and L.) Thus, one concludes that equation (34) is valid for $\hat{b} \neq 0$.

As an example, the existence of the integers K_o and L_o referred to in (3) will be established as follows for the point γ_1 with the assumption $\hat{b}_1 > 0$.

Proof

Since $(s_x)^2 + (s_y)^2 = (s_1^+)^2 + (s_2^+)^2 = 1$, it need only be proven that, given any $\epsilon_x > 0$, $\epsilon_y > 0$, and $\delta_x > 0$, there exist K_o and L_o such that for any $K \geq K_o$ and $L \geq L_o$, two integers k $(K > k > 0)$ and ℓ $(L > \ell > 0)$ can be found to satisfy the following conditions:

$$\left.\begin{aligned} \delta_x &> |s_x - s_1^+| \\ \epsilon_x &> |t_x - 1| \\ \epsilon_y &> \left|t_y - \frac{\hat{b}}{|\hat{b}|}\right| \end{aligned}\right\} \tag{B21}$$

To proceed, note that, without any loss of generality, it may be assumed that

$$s_1^+ > \delta_x \qquad 1 - s_1^+ > \delta_x \tag{B22}$$

With the aid of the assumption $\hat{b} > 0$ and expressions (B14), (B16), (B17), and (B22), equation (B21) can be rewritten as

$$\left.\begin{aligned} \eta^+ &> \frac{\sin(\pi\ell/L)}{\sin(\pi k/K)} > \eta^- \\ \epsilon_x &> 2\sin^2\left(\frac{\pi k}{2K}\right) \\ \epsilon_y &> 2\sin^2\left(\frac{\pi\ell}{2L}\right) \end{aligned}\right\} \tag{B23}$$

where

$$\eta^+ \stackrel{\text{def.}}{=} \frac{\Delta y}{\Delta x}\sqrt{\frac{a_o}{c_o}}\left[\frac{\sqrt{1-(s_1^+ - \delta_x)^2}}{s_1^+ - \delta_x}\right] > \eta^-$$

$$\eta^- \stackrel{\text{def.}}{=} \frac{\Delta y}{\Delta x}\sqrt{\frac{a_o}{c_o}}\left[\frac{\sqrt{1-(s_1^+ + \delta_x)^2}}{s_1^+ + \delta_x}\right] > 0$$

It is easy to show that there exist two integers $k_o > 1$ and $\ell_o > 1$ such that

$$\eta^+ > \frac{\ell_o + 1}{k_o - 1} > \frac{\ell_o - 1}{k_o + 1} > \eta^- \tag{B24}$$

Since

$$\lim_{z \to 0} \frac{\sin[(\ell_o + 1)z]}{\sin[(k_o - 1)z]} = \frac{\ell_o + 1}{k_o - 1}$$

$$\lim_{z \to 0} \frac{\sin[(\ell_o - 1)z]}{\sin[(k_o + 1)z]} = \frac{\ell_o - 1}{k_o + 1}$$

One can find an integer $M \gg \text{Max}\ \{k_o,\ell_o\}$ such that

$$\eta^+ > \frac{\sin\ [\pi(\ell_o + 1)/M]}{\sin\ [\pi(k_o - 1)/M]} > \frac{\sin\ [\pi(\ell_o - 1)/M]}{\sin\ [\pi(k_o + 1)/M]} > \eta^- \tag{B25}$$

and

$$\left.\begin{aligned} \epsilon_x &> 2\ \sin^2\left[\frac{\pi(k_o + 1)}{2M}\right] \\ \epsilon_y &> 2\ \sin^2\left[\frac{\pi(\ell_o + 1)}{2M}\right] \end{aligned}\right\} \tag{B26}$$

Let

$$\left.\begin{aligned} K_o &\overset{\text{def.}}{=} Mk_o \\ L_o &\overset{\text{def.}}{=} M\ell_o \end{aligned}\right\} \tag{B27}$$

For any $K \geq K_o$ and $L \geq L_o$, there exist six unique integers α, β, γ, α', β', and γ' such that

$$\left.\begin{aligned} K &= \alpha M k_o + \beta M + \gamma \\ L &= \alpha' M \ell_o + \beta' M + \gamma' \end{aligned}\right\} \tag{B28}$$

and

$$\left.\begin{aligned} &\alpha \geq 1 \quad k_o > \beta \geq 0 \quad M > \gamma \geq 0 \\ &\alpha' \geq 1 \quad \ell_o > \beta' \geq 0 \quad M > \gamma' \geq 0 \end{aligned}\right\} \tag{B29}$$

If

$$\left.\begin{aligned} k &\overset{\text{def.}}{=} k_o(\alpha k_o + \beta) \\ \ell &\overset{\text{def.}}{=} \ell_o(\alpha' \ell_o + \beta') \end{aligned}\right\} \tag{B30}$$

are chosen, with the aid of expressions (B28) and (B29), it can be shown that

$$\left.\begin{aligned} \frac{k_o + 1}{M} > \frac{k}{K} > \frac{k_o - 1}{M} \\ \frac{\ell_o + 1}{M} > \frac{\ell}{L} > \frac{\ell_o - 1}{M} \end{aligned}\right\} \tag{B31}$$

Using inequalities (B25), (B26), and (B31), and the fact that $M \gg \text{Max}\ \{k_o,\ell_o\}$, it can be shown that inequality (B23) along with the conditions $K > k > 0$ and $L > \ell > 0$ are satisfied if k and ℓ are chosen according to equation (B30). QED

This appendix concludes with a discussion on expressions (B19) and (B20). With the aid of equations (27) to (29) and (35) to (37), equation (B19) implies that $\tau^* = \tau^o$ and $G^* = G^o$ when $\hat{b} = 0$ even if the integers K and L are finite. On the other hand, for $\hat{b} \neq 0$, inequality (B20) implies that (1) $G^* > G^o$ always and (2) $\tau^o \neq \tau^*$ unless $\gamma_{\max} + \gamma_{\min} = \lambda_{\max} + \lambda_{\min}$. Since the current local relaxation procedure and convergence rate prediction method are developed from the assumptions that $G^* = G^o$ and $\tau^* = \tau^o$, one may expect that the current procedure works less well, and the predictions given by the parameter $O_t(n)$ become more conservative for a PDE with a cross-derivative term. The second part of the above expectations was confirmed by the numerical results shown in the section Numerical Evaluation.

Appendix C
Derivation of Equations (39) to (41)

With the aid of equations (30) to (32), equation (37) can be rewritten as

$$\Sigma^* = \frac{a\alpha_o + c + \sqrt{(a\alpha_o - c)^2 + 4b^2\alpha_o}}{a\alpha_o + c - \sqrt{(a\alpha_o - c)^2 + 4b^2\alpha_o}} \geq 1 \qquad \text{(C1)}$$

where $\alpha_o \overset{\text{def.}}{=} c_o/a_o > 0$. It should be noted that, as a result of equation (4), the denominator of the fraction in equation (C1) is always positive. In view of equations (36) and (C1), one may consider the parameter G^* as a function of a, b, c, and α_o, and obtains

$$\frac{\partial G^*}{\partial \alpha_o} = \frac{8a(ac-b^2)(\alpha_o - c/a)}{(\Sigma^*+1)^2 \cdot \sqrt{(a\alpha_o - c)^2 + 4b^2\alpha_o} \cdot \left[a\alpha_o + c - \sqrt{(a\alpha_o - c)^2 + 4b^2\alpha_o}\right]^2} \qquad \text{(C2)}$$

where $(a\alpha_o - c)^2 + 4b^2\alpha_o \neq 0$ is assumed. Equation (C2) coupled with equation (4) implies that (1) $\partial G^*/\partial \alpha_o < 0$ if $\alpha_o < c/a$, and (2) $\partial G^*/\partial \alpha_o > 0$ if $\alpha_o > c/a$. Equations (39) and (40) are the direct results of properties (1) and (2).

To prove equation (41), one notes that $\hat{a} = \hat{c}$ if $c_o/a_o = c/a$. Thus equations (B13) to (B17) imply that

(1) $\gamma^{(k,\ell)} + \gamma^{(K-k,\ell)} = \gamma^{(k,\ell)} + \gamma^{(k,L-\ell)} = 2\hat{a}$

and

(2) $\gamma^{(0,\ell)} = \gamma^{(k,0)} = \hat{a}$

for $k = 1,2,\ldots,(K-1)$ and $\ell = 1,2,\ldots,(L-1)$. Consequently one concludes that

$$\gamma_{\max} + \gamma_{\min} = \lambda_{\max} + \lambda_{\min} = 2\hat{a} \qquad \text{(C3)}$$

Equation (41) follows directly from equations (28), (35), and (C3).

Appendix D
Derivation of Equations (53) and (54)

To obtain the results given in equations (53) and (54), the parameters $\alpha_o \stackrel{\text{def.}}{=} c_o/a_o > 0$ and $\beta_{ij} \stackrel{\text{def.}}{=} c_{ij}/a_{ij} > 0$ are introduced. With the assumption that $b_{ij} = 0$ for all $(i,j) \in \Phi$, equations (30) to (32), (37), and (49) can be used to show that

$$G_{ij} = J(\alpha_o/\beta_{ij}) \stackrel{\text{def.}}{=} \begin{cases} \dfrac{(\alpha_o/\beta_{ij}) - 1}{(\alpha_o/\beta_{ij}) + 1} & \text{if } \alpha_o/\beta_{ij} \geq 1 \\ \dfrac{1 - (\alpha_o/\beta_{ij})}{(\alpha_o/\beta_{ij}) + 1} & \text{if } 1 \geq \alpha_o/\beta_{ij} > 0 \end{cases} \tag{D1}$$

Note that

$$\frac{dJ(t)}{dt} = \begin{cases} \dfrac{2}{(t+1)^2} > 0 & \text{if } t > 1 \\ \dfrac{-2}{(t+1)^2} < 0 & \text{if } 1 > t > 0 \end{cases} \tag{D2}$$

That is, the function $J(t)$ increases monotonically if $t > 1$ and decreases monotonically if $1 > t > 0$.

Let $\beta_{\max}$ and $\beta_{\min}$ be the parameters defined in equation (55) and, for a given α_o,

$$\hat{J}(\alpha_o) \stackrel{\text{def.}}{=} \underset{(i,j)\in\Phi}{\text{Max}} \{J(\alpha_o/\beta_{ij})\} \tag{D3}$$

Assuming $\beta_{\max} > \beta_{\min}$, it can be shown that

$$\left.\begin{array}{ll} \dfrac{d\hat{J}(\alpha_o)}{d\alpha_o} > 0 & \text{if } \alpha_o > \alpha_m \\ \dfrac{d\hat{J}(\alpha_o)}{d\alpha_o} < 0 & \text{if } \alpha_o < \alpha_m \end{array}\right\} \tag{D4}$$

where $\alpha_m \stackrel{\text{def.}}{=} \sqrt{\beta_{\max} \cdot \beta_{\min}}$

Proof

As a result of equation (D2), $\hat{J}(\alpha_o)$ equals to the greater of $J(\alpha_o/\beta_{\min})$ and $J(\alpha_o/\beta_{\max})$. Since (1) $\alpha_m/\beta_{\min} > 1$, (2) $\alpha_m/\beta_{\max} < 1$, and (3) $J(\alpha_m/\beta_{\min}) = J(\alpha_m/\beta_{\max})$, one concludes that

$$\hat{J}(\alpha_o) = \left\{\begin{array}{ll} J(\alpha_o/\beta_{\min}) & \text{if } \alpha_o \geq \alpha_m \\ J(\alpha_o/\beta_{\max}) & \text{if } \alpha_o \leq \alpha_m \end{array}\right\} \tag{D5}$$

Inequality (D4) is a direct result of equations (D2) and (D5). QED

Inequality (D4) implies that $\hat{J}(\alpha_o)$ increases monotonically if $\alpha_o > \alpha_m$ and decreases monotonically if $\alpha_o < \alpha_m$. Since $\hat{J}(\alpha_o) = G^{\infty}$ (eqs. (51), (D1), and (D3)), equations (53) and (54) simply state the fact that $\hat{J}(\alpha_o)$ reaches its minimum $\hat{J}(\alpha_m) = J(\alpha_m/\beta_{\min}) = G^{\infty}_{\min}$ if and only if $\alpha_o = \alpha_m$.

In case that $\beta_{\max} = \beta_{\min}$, equations (D1) and (D3) imply that the minimum of $\hat{J}(\alpha_o)$ is zero and it is reached if and only if $\alpha_o = \beta_o$ where β_o denotes the value of either $\beta_{\max}$ or $\beta_{\min}$. Equations (53) and (54) obviously are valid for this special case.

References

1. Hockney, R.W.: Fast Direct Solution of Poisson's Equation using Fourier Analysis. J. Assoc. Comput. Mach., vol. 12, no. 1, Jan. 1965, pp. 95–113.
2. Hockney, R.W.: The Potential Calculation and Some Applications. Methods in Computational Physics, Vol. 9, Plasma Physics, B. Adler, S. Fernbach and M. Rotenberg, eds., Academic Press, 1970, pp. 135–211.
3. Dorr, F.W.: The Direct Solution of the Discrete Poisson Equation on a Rectangle. SIAM Rev., vol. 12, no. 2, Apr. 1970, pp. 248–263.
4. Buneman, O.: A Compact Non-iterative Poisson Solver. Institute for Plasma Research, Stanford University, Report SU-IPR-294, May 1969.
5. Hockney, R.W.: Rapid Elliptic Solvers. Numerical Methods in Applied Fluid Dynamics, B. Hunt, ed., Academic Press, 1980, pp. 1–48.
6. D'Yakanov, E.G.: An Iteration Method of Solving Simultaneous Equations of Finite Differences. Dokl. Akad. Nauk. SSSR, vol. 138, 1961, pp. 522–525.
7. Concus, P.; and Golub, G.H.: Use of Fast Direct Methods for Efficient Numerical-Solution of Non-Separable Elliptic Equations. SIAM J. Numer. Anal., vol. 10, no. 6, Dec. 1973, pp. 1103–1119.
8. Bank, R.E.: Marching Algorithms for Elliptic Boundary-Value Problems. II —The Variable Coefficient Case. SIAM J. Numer. Anal., vol. 14, no. 5, Oct. 1977, pp. 950–970.
9. Chang, S.C.; and Adamczyk, J.J.: A New Approach for Solving the Three-Dimensional Steady Euler Equations: Part I—General Theory, and, Part II—Application to Secondary Flows in a Turning Channel. J. Comput. Phys., vol. 60, no. 1, Aug. 1985, pp. 23–61.
10. Botta, E.F.F.; and Veldman, A.E.P.: On Local Relaxation Methods and Their Application to Convection-Diffusion Equations. J. Comput. Phys., vol. 48, no. 1, Oct. 1982, pp. 127–149.
11. Chang, S.C.: A Semi-Direct Procedure Using a Local Relaxation Factor and its Application to an Internal Flow Problem. Ninth International Conference on Numerical Methods in Fluid Dynamics. Soubbaramayer; and J.P. Boujot, eds., Springer-Verlag, 1984, pp. 143–147.
12. Smith, Gordon D.: Numerical Solution of Partial Differential Equations, Second ed., Clarendon Press, Oxford, 1978, p. 29.
13. Hageman, L.A.; and Young, D.M.: Applied Iterative Methods. Academic Press, 1981, p. 12.
14. Adams, J.; Swarztrauber, P.; and Sweet, R.: FISHPAK: A Package of FORTRAN Subprograms for the Solution of Separable Elliptic Partial Differential Equations, Version 3. National Center for Atmospheric Research, Boulder, CO, 1979.
15. Chang, S.C.: Solution of Elliptic Partial Differential Equations by Fast Poisson Solvers Using A Local Relaxation Factor II—Two-step Method. NASA TP-2530, 1986.
16. Chang, S.C.: Generalizations of Two Inequalities Involving Hermitian Forms. Linear Algebra Appl., vol. 65, Feb. 1985, pp. 179–187.

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
NASA TP-2529		

4. Title and Subtitle	5. Report Date
Solution of Elliptic Partial Differential Equations by Fast Poisson Solvers Using a Local Relaxation Factor I - One-Step Method	May 1986
	6. Performing Organization Code 505-31-04
7. Author(s) Sin-Chung Chang	8. Performing Organization Report No. E-2461-1
	10. Work Unit No.
9. Performing Organization Name and Address National Aeronautics and Space Administration Lewis Research Center Cleveland, Ohio 44135	11. Contract or Grant No.
	13. Type of Report and Period Covered Technical Paper
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Washington, D.C. 20546	14. Sponsoring Agency Code

15. Supplementary Notes

Presented in part at Ninth International Conference on Numerical Methods in Fluid Dynamics, Saclay, France, June 25-29, 1984.

16. Abstract

An algorithm for solving a large class of two- and three-dimensional nonseparable elliptic partial differential equations (PDE's) is developed and tested. It uses a modified D'Yakanov-Gunn iterative procedure in which the relaxation factor is grid-point dependent. It is easy to implement and applicable to a variety of boundary conditions. It is also computationally efficient, as indicated by the results of numerical comparisons with other established methods. Furthermore the current algorithm has the advantage of possessing two important properties which the traditional iterative methods lack; that is, (1) the convergence rate is relatively insensitive to grid-cell size and aspect ratio, and (2) the convergence rate can be easily estimated by using the coefficient of the PDE being solved.

17. Key Words (Suggested by Author(s))	18. Distribution Statement
Elliptic problem solver One-step local relaxation method	Unclassified - unlimited STAR Category 64

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of pages	22. Price*
Unclassified	Unclassified	21	A02

*For sale by the National Technical Information Service, Springfield, Virginia 22161

NASA-Langley, 1986

National Aeronautics and
Space Administration
Code NIT-4

Washington, D.C.
20546-0001

BULK RATE
POSTAGE & FEES PAID
NASA
Permit No. G-27

POSTMASTER: **If Undeliverable (Section 158 Postal Manual) Do Not Return**